

South Dakota AFNR

Academic Integration Activities: Example #14

→ *Fundamental Plant Science students complete soil tests, examine the data, and draw conclusions.*

1. Ag Standard

Fundamental Plant Science PS.2.2

Students examine data to evaluate and manage soil/media and nutrients.

- > Test soil/media and plant tissue for nutrient levels.
- > Interpret test of soil/media and/or plant tissue.

2. Academic Standard

9-12.S.1.1

Students are able to draw conclusions from a set of data.

3. Background Information

Five-Number Summary

Consists of the upper and lower extremes, the median, and the upper and lower quartile.

- > Upper extreme = Greatest number that occurs in the set
- > Lower extreme = Lowest number that occurs in the set
- > Median = Middle term when data arranged least to greatest
- > Upper quartile = Median of the upper half of data
- > Lower quartile = Median of the lower half of data

Steps to Determining Quartiles

- a. Arrange data in order from least to greatest.
- b. Identify the median.
- c. Identify the middle term of each half of the data on either side of the median. These values are the upper and lower quartiles.

Example

Consider the set {1, 3, 4, 5, 6, 7, 9}.

- > The lower extreme is 1.
- > The lower half is {1, 3, 4}, and the middle term of that half is 3. Therefore, the lower quartile is 3.
- > The median is the middle term, 5.
- > The upper half is {6, 7, 9}, and the middle term of that half is 7. Therefore, the upper quartile is 7.
- > The upper extreme is 9.

Outliers

An outlier is a value in a data set that is far from the other values. Outliers can be caused by experimental errors, measurement errors, or by a long-tailed population. In the former cases, it can be desirable to identify outliers and remove them from data before performing a statistical analysis,

because they can throw off the results and inaccurately represent the sample population. The simplest way to identify outliers is with the quartile method.

Instructions for Finding Outliers

- a. Sort the data in ascending order.
 - > Example data set: {4, 5, 2, 3, 15, 3, 3, 5}.
 - > Sorted, it is {2, 3, 3, 3, 4, 5, 5, 15}.
- b. Find the median. This is the number at which half of the data points are larger and half are smaller. If there is an even number of data points, the middle two are averaged.
 - > In the example data set, the middle points are 3 and 4, so the median is $(3 + 4)/2$, or 3.5.
- c. Find the upper quartile, Q2. This is the data point at which 25% of the data are larger. If the data set is even, average the two points around the quartile.
 - > In the example data set, this is $(5 + 5)/2$, or 5.
- d. Find the lower quartile, Q1. This is the data point at which 25% of the data are smaller. If the data set is even, average the two points around the quartile.
 - > In the example data set, $(3 + 3)/2$, or 3.
- e. Subtract the lower quartile from the higher quartile to get the interquartile range, IQ.
 - > In the example data set, $Q2 - Q1 = 5 - 3$, or 2.
- f. Multiply the interquartile range by 1.5. Add this to the upper quartile and subtract it from the lower quartile. Any data point outside these values is a mild outlier.
 - > In the example set, $1.5 \times 2 = 3$. $3 - 3 = 0$ and $5 + 3 = 8$. So any value less than 0 or greater than 8 would be a mild outlier. This means that 15 qualifies as an outlier.

4. Example in Context

Students test the soil pH of their home gardens. Each student brings in a sample of the soil used in their home gardens and runs pH tests. The class data is statistically analyzed.

What is the five-number summary of the data set?

Is there an outlier in the data set?

Review how to determine the five-number summary and find an outlier.

Five-Number Summary

- a. Arrange data in ascending order.
- b. Record lower extreme. (lowest value)
- c. Record higher extreme. (highest value)
- d. Record median. (middle term)
- e. Record lower quartile. (median of lower half of data)
- f. Record upper quartile. (median of upper half of data)

Outlier

- a. Arrange data in ascending order.
- b. Find median.
- c. Find upper quartile.
- d. Find lower quartile.
- e. Subtract the lower quartile from the higher quartile to get the interquartile range.
- f. Multiply the interquartile range by 1.5. Add this to the upper quartile and subtract from the lower quartile. Any data outside of these values is an outlier.

The following is the pH data from the students' home gardens:
6.7, 6.8, 6.5, 6.9, 6.2, 7.2, 6.4, 6.2, 6.6

What is the five-number summary of the data set?

Arrange data in ascending order.

6.2, 6.2, 6.4, 6.5, 6.6, 6.7, 6.8, 6.9, 7.2

Record lower extreme = 6.2

Record higher extreme = 7.2

Record median = 6.6

Record lower quartile = 6.3

Record upper quartile = 6.85

Answer: Five-number summary = (6.2, 6.3, 6.6, 6.85, 7.2)

Is there an outlier in the data set?

Arrange data in ascending order.

6.2, 6.2, 6.4, 6.5, 6.6, 6.7, 6.8, 6.9, 7.2

Find median = 6.6

Find lower quartile = 6.3

Find upper quartile = 6.85

Subtract the lower quartile from the higher quartile. Interquartile range: $6.85 - 6.3 = 0.55$

Multiply interquartile range by 1.5. $0.55 \times 1.5 = 0.825$

Add that to the upper quartile. $6.85 + 0.825 = 7.68$

Subtract that from the lower quartile. $6.3 - 0.825 = 5.48$

Answer: There are no values in this data set that are outside the range of 5.48 to 7.68. Therefore, no, there are not any outliers in this data set.

5. Guided Practice Exercise

Students test the various soil nitrate levels using testing strips.



The following is the amount of nitrate found in each soil test:

45 ppm, 30 ppm, 20 ppm, 25 ppm, 5 ppm, 35 ppm, 40 ppm, 45 ppm, 40 ppm, 20 ppm, 30 ppm

What is the five-number summary of the data set?

Arrange data in ascending order.

5, 20, 20, 25, 30, 30, 35, 40, 40, 45, 45

Record lower extreme = 5

Record higher extreme = 45

Record median = 30

Record lower quartile = 20

Record upper quartile = 40

Answer: Five-number summary = (5, 20, 30, 40, 45)

Is there an outlier in the data set?

Arrange data in ascending order.

5, 20, 20, 25, 30, 30, 35, 40, 40, 45, 45

Find median = 30

Find lower quartile = 20

Find upper quartile = 40

Subtract the lower quartile from the higher quartile. Interquartile range = 20

Multiply interquartile range by 1.5. $20 \times 1.5 = 30$

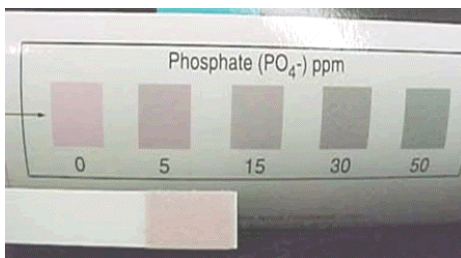
Add that to the upper quartile. $40 + 30 = 70$

Subtract that from the lower quartile. $20 - 20 = 0$

Answer: There are no values in this data set that are outside the range of 0 to 70. Therefore, no, there are not any outliers in this data set.

6. Independent Practice Exercises

Students test the various soil phosphate levels using testing strips.



The following is the amount of nitrate found in each soil test:

5 ppm, 20 ppm, 15 ppm, 0 ppm, 25 ppm, 40 ppm, 20 ppm, 15 ppm, 5 ppm, 20 ppm, 30 ppm

What is the five-number summary of the data set?

Answer: Five-number summary = (0, 5, 20, 25, 40)

Arrange data in ascending order: 0, 5, 5, 15, 15, 20, 20, 20, 25, 30, 40

Record lower extreme = 0
Record higher extreme = 40
Record median = 20
Record lower quartile = 5
Record upper quartile = 25

Is there an outlier in the data set?

Answer: No, there are not any values outside of -25 to 55, so there are no outliers in this data set.

Arrange data in ascending order: 0, 5, 5, 15, 15, 20, 20, 20, 25, 30, 40
Find median = 20
Find lower quartile = 5
Find upper quartile = 25
Subtract the lower quartile from the higher quartile. Interquartile range = 20
Multiply interquartile range by 1.5. $20 \times 1.5 = 30$
Add that to the upper quartile. $25 + 30 = 55$
Subtract that from the lower quartile. $5 - 30 = -25$

Students test the soil pH of the soils around the school building prior to planting flowers. Here is the recorded pH data from various locations and a representative sample:

5.8, 5.1, 5.0, 8.0, 5.3, 5.5, 5.8

What is the five-number summary of the data set?

Answer: (5.0, 5.1, 5.5, 5.8, 8.0)

Is there an outlier in the data set?

Answer: Yes, 8.0 is an outlier because it is out of the acceptable range of 4.05 to 6.85.

7. Notes

[None]